

ADAPTIVE MULTI-RESOLUTION SCHEME FOR EFFICIENT IMAGE COMPRESSION

Babel M., Déforges O., Ronsin J.

CNRS UMR 6164 IETR laboratory / INSA Rennes
20, av des Buttes de Coëmes, CS 14315, 35043 Rennes Cedex, France
Contact : {odeforge, mbabel, jronsin}@insa-rennes.fr

ABSTRACT

The LAR (Locally Adaptive Resolution) method is a multi-layers still image coding scheme, efficient from very low to high bit rates. The first stage is devoted to the representation and compression of the global information (low resolution image), and relies on an adaptive resolution in the image. This paper presents some improvements on the first layer through an original quad-tree like decomposition based on a predictive scheme, and the integration of a powerful interpolation post-processing. Comparative results with state of the art compression method are also presented and discussed.

1. INTRODUCTION

Recent developments in image coding have demonstrated that H.264 appears as the most efficient solution for both video and still image compression [1]. In particular, it can over-class Jpeg2000 at every bit-rates but at the depend of computation time [2]. However, H.264 does not provide a scalable bitstream.

The LAR (Locally Adaptive Resolution) method has been originally introduced as a fast method for high compression of gray-level images [3]. Extensions have permitted to address also middle and high quality encoding, automatic reconstruction of a region representation at the decoder allowing both Region Of Interest coding and region-level chromatic images encoding [4].

This paper is organized as follows : section 2 recalls the outlines of the LAR method, while section 3 presents the new advances over the first high compression stage. Some comparative results are given in section 4, and we conclude in section 5.

2. LAR METHOD OVERVIEW

One of the main improvements introduced in H.264 is the adaptation to the image content by considering different sizes of block. For Intra-picture mode, flat areas are encoded using 16x16 blocks, whereas more active ones involve 4x4 blocks. Our coding scheme is based on the same principle but declined much further. Indeed, the

block size fits the actual local resolution so that a meaningful low resolution image is provided.

The LAR method is a two layers codec (Fig. 1) : a spatial and a spectral one. Low bit rate compression can be achieved when considering only the spatial coder. The second layer allows encoding the image texture.

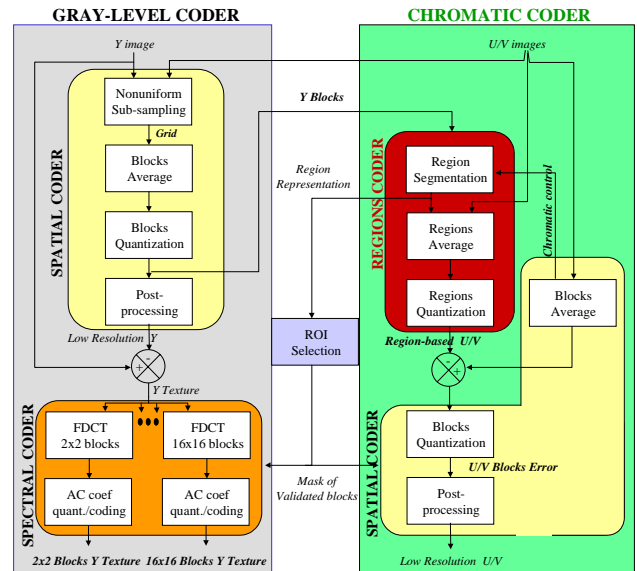


Figure 1: Global LAR codec scheme

2.1 Spatial Coder

The principle relies on a variable image resolution. The image is divided into macroblocks (16x16) possibly split up from a local activity estimation onto a quadtree data structure until 2x2 size. A low resolution image can be obtained when filling each block by its average intensity. Many works use such representation, but our approach differs by three essential points:

- the activity is estimated from a local morphological gradient. As a result, the variable block size representation constitutes a coarse edge-driven segmentation map (small blocks upon contours, large blocks in smooth areas).
- simple linear quantizations can be applied but based on psychovisual concepts : fine quantization for large blocks (strong sensitivity of human eye to brightness

variations in uniform areas), coarse quantization for small blocks (lower sensitivity).

A classical quad-tree decomposition would lead to an expensive encoding of the block luminance value. Till now, the original LAR method was based on a one raster scan DPCM scheme. The next section will introduce a new multi-resolution decomposition providing both efficiency and scalability.

The output of the spatial coder is an image subsampled by two preserving edges (2x2 blocks transformed into single pixel). Perceptible block artifacts in homogeneous areas are easily removed by a simple efficient post-processing based on an adaptive bilinear interpolation (Fig. 2).

We can summarize the major features of this coder:

- a very fast and efficient technique for high compression ratio,
- a method simplifying the source image by removing the local texture but globally preserving the object boundaries.
- a method naturally allowing a spatial scalability (CIF -> QCIF for instance), suitable for broadcast applications involving different kind of receivers [5].

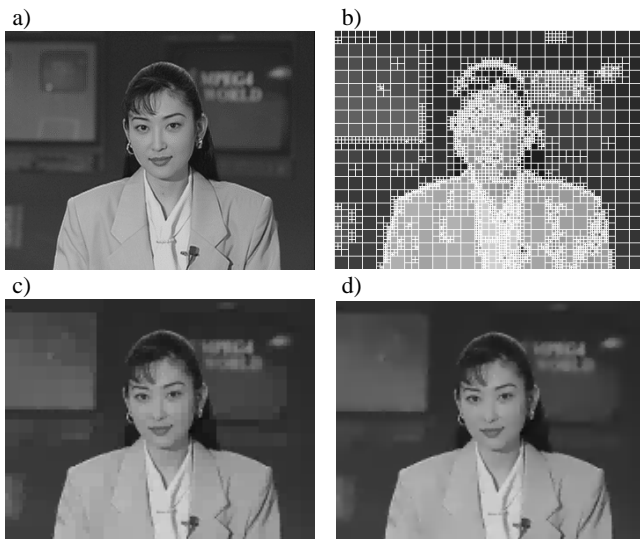


Figure 2: Spatial codec,

a) Source image (CIF), b) variable size grid : 0.0029 bpp, c) Var. bl. size image : 0.123 bpp (QCIF), d) after postprocess.

2.2 Spectral coder

To obtain higher image quality a second spectral coder can be added to encode the whole error image (texture image). We use a DCT adaptive block-size approach, but where both the size and the DC components are provided by the spatial coder. The use of adapted square sizes offers two major advantages :

- The dynamic of the block content is bounded (except for 2x2 blocks). So fewer AC coefficients are generated with lower energy than for standard fixed block size DCT.

- As the block sizes are coded independently the coding scheme is content-based scalable (small blocks -> edge enhancement, large blocks -> smooth area refinement).

This second coder can still be improved by including some H.264 techniques such as spatial AC prediction, two zigzag scans within 4x4 blocks, and context modeling together with the entropy coding.

2.3 self-extracting region representations

Unlike traditional compression techniques, the low resolution LAR image does not introduce strong distortions which usually prevent from a reliable segmentation process. Indeed, the variable block size representation resulting from a local gradient distance can constitute a "split" step in a segmentation scheme. The original idea of the self-extracting region representation is to transmit the Y low resolution LAR image, and to continue the segmentation ("merge" step) at both the sender and receiver. As regions are built as union of blocks, two main advantages are introduced by the method :

- a free region representation (no cost overhead for the contours description),
- contours regions and texture coding share the same grid of representation.

A Region Of Interest can then be selected either at the coder or the decoder, and the spectral coder activated only for blocks included in this ROI.

By adding a very low cost chromatic control during the segmentation process, reliable regions in terms of regions are generated, allowing very high compression rate coding of U and V images at the region levels [4]. Coding the error images at the block level (spatial coder) leads to very good quality chromatic images.

3. SPATIAL CODER IMPROVEMENT

3.1 Pyramidal predictive decomposition

To get a scalable low resolution image while maintaining a compression efficiency, we have introduced a new quad-tree decomposition scheme. The previous spatial LAR coder encoded in one pass (raster scan) the intensity block values by means of the Gradient Adjusted Predictor (GAP) [6]. As for the new scalable DPCM scheme, it is based on the predictor described in [7] by Wu. For full resolution image, errors are coded by means of three interlaced sampling of the original image. By this way, we tend to obtain a spatial configuration of 360° type context surrounding a given pixel.

The general principle is as follows (Fig. 3) : the first pass encodes through a classical DPCM system an uniform subsampled image formed by the average of two diagonally adjacent pixels within each 2x2 block. Then the second pass will provide the values of the two pixels

used to compute the previous image. At this stage, the prediction 360° degrees type context consists in the already known values of the current pass and the diagonal means coded by the first pass. Finally, the third pass encodes the remaining half of the original image. Once again, thanks to the reconstructed pixels, a completely spatially enclosing and adjacent context is available to predict the modeled pixel.

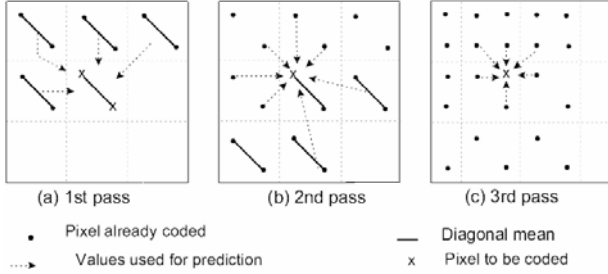


Figure 3: Wu algorithm : 3 passes, prediction neighborhood.

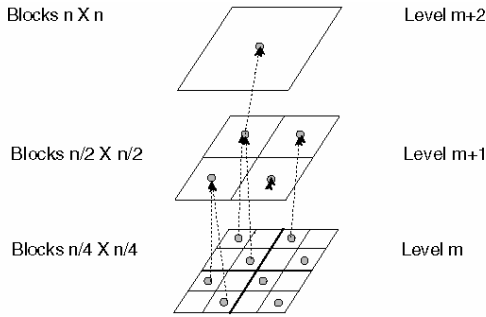


Figure 4: Bottom-up building

We propose to extend the previous full resolution scanning scheme to get a pyramidal representation of the image.

Bottom-Up building

Starting from the initial LAR partition, four blocks $N/2 \times N/2$ are gathered into one block $N \times N$ valued by the average of the two blocks of the first diagonal (Fig. 4). By this way, the value of a $N \times N$ block is equal to the average over the whole diagonal in the finest resolution image. The highest level (level 4) of the pyramid corresponds to the 16×16 blocks (largest block size in the LAR method).

Top-Down decomposition

Let the original image be of width W and height H , the first stage of our pyramid concerns the image of size $W/16 \times H/16$ (level 4). The first pass of the Wu algorithm (simple causal DPCM coding of the block values) is applied. Then, each square is split until its maximal size is reached. For a given level (block size) of the pyramid, the

blocks of smaller or equal size are processed by successively passes 2 and 3. The values of higher size blocks are copied out in order to refine the context (Fig.5).

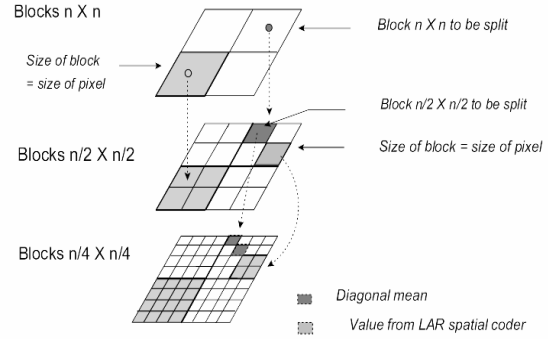


Figure 5: Top-down decomposition

By this way, taking into account the variable resolution LAR image, a progressive content-based method is developed : if a restricted context is sufficient for the large blocks, a more accurate and informative context exists for the small blocks (objects boundaries).

The main property of this method is that the pyramidal decomposition does not introduce additional symbols when applying a minimal quantization : the number of encoded symbol is equal to the number of blocks of our LAR representation. Indeed, for a group of four blocks obtained by the split of a block from the above level, the diagonal mean is already known. The exact value of the second value on the diagonal can be deduced with one bit knowing the first value. This bit disappears with quantization by a minimum step of two. Figure 6 gives a representation of the decomposition. This method has also been extended to lossless compression, nearly achieving performances of state of the art CALIC [7], while offering a scalable coding scheme.

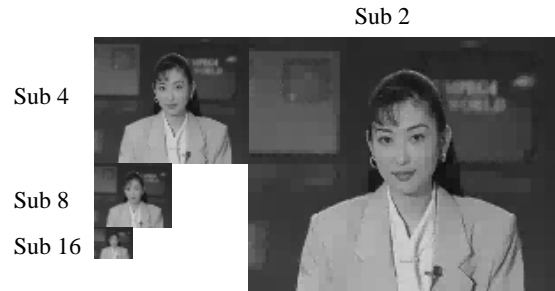


Figure 6: Multiresolution representation : grid : 0.029 bpp, lev1 (sub 16) : 0.005 bpp, lev2 (sub 8) : 0.015 bpp, lev3 (sub 4) : 0.051 bpp, lev4 (sub 2) : 0.094 bpp

3.2 Interpolation

From our low resolution image sub-sampled by two, an interpolation stage can be envisaged in order to build a full resolution image.

We have then introduced the very efficient adaptive image interpolation described in [8], based on the theory of optimal recovery. After determining locally, at a given location, the quadratic signal class from a set of training vectors, optimal recovery estimates the missing pixels. By this way, designing an interpolation filter is equivalent to minimize the integrated square error in the frequency domain. This new type of interpolation is essentially based on learning : the coarse scale image can be then used to improve the resulting image.

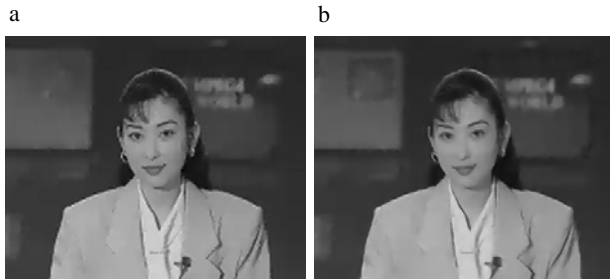


Figure 7: Full size (0.123 bpp), a) bi-cubic b) optimal recovery

Hence, we have applied this algorithm on the subsampled post-processed luminance image provided by the LAR spatial coder. Clearly, images obtained thanks to this method are of visually much better quality than traditional technique such as bi-cubic interpolation, especially on the contours (no more jagged effect). Computation time is nevertheless more prohibitive (about one second on a standard PC).

Comparative results are given figure 8 between intra mode Mpeg-4 (Momusys), Jpeg2000, intra mode H.261 (most optimized mode) and the LAR approach. Jpeg2000 is a fully scalable coding method, the LAR is partially scalable (by layers), whereas Mpeg4 and H.261 are not scalable. While LAR's PSNR is about 1.5 db lower than H.261, the LAR method does not introduce the same kind of distortions as the other approaches, smoothes texture details, but preserves a pleasant visual aspect.

5. CONCLUSION AND PERSPECTIVES

We have introduced new improvements realized for the LAR codec. The main idea is to build a totally scalable algorithm, where the low resolution image is of good quality. Even if H26L uses also a decomposition with different but fewer block sizes, progressivity is not one of its properties. Visually one can observe the good performances of the LAR coding given thanks to the optimal recovery interpolation.

Further works will concern the lowering of the cost of the AC coefficients at the spectral coder : H26L proposes

efficient schemes in this domain. Thereby, LAR coder could be seen as a new functionality of this state-of-the-art.

6. REFERENCES

- [1] G. Côté, L. Winger, "Recent Advances in Image Compression Standards", IEEE Canadian Review on Computers, spring 2002, pp 20-24.
- [2] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Document: JVT-D039, "Performance comparison: H.26L intra coding vs. JPEG2000", July 2002, pp 1-9.
- [3] O. Déforges and J. Ronsin, "Nonuniform sub-sampling using squares elements : a fast still image coding at low bit rate", *International PCS'99*, Portland, 21-23 april 1999.
- [4] O. Déforges, J. Ronsin, "Region of Interest Coding for Low Bit-Rate Image Transmission", *ICME'2000*, New-York, USA, 31 July -02 August 2000.
- [5] H. Sun, W. Kwork, J.W. Zdepski, "Architectures for MPEG compressed domain bitstream", *IEEE Trans. on Circuits Systems Video techno.*, vol 6, pp 191-199, apr. 1996.
- [6] X. Wu, N. Memon, and K. Sayood, "A context-based, adaptive, lossless/nearly-lossless coding scheme for continuous-tone images," in *ISO/IEC JTC 1/SC 29/WG 1*, vol. 202, July 1995.
- [7] X. Wu, "Lossless compression of Continuous-Tone Images via Context Selection, Quantization and Modelling", *IEEE Trans. On Image Processing*, vol. 6, no. 5, pp. 656-664, 1996.
- [8] D. D. Muresan and T. W. Parks, "Optimal Recovery Approach to Image Interpolation", *IEEE ICIP 2001*, pp. Image Interpolation and Spatial Filtering, Greece, 2001.



Figure 8: Comparative results (0.123 bpp), a) Mpeg4 b) jpeg2000, c) H.261, d) LAR